

```
!pip install reportlab
!pip install pdfminer.six
```

```
Collecting reportlab
```

```
  Downloading reportlab-4.1.0-py3-none-any.whl (1.9 MB)
```

```
1.9/1.9 MB 7.0 MB/s eta 0:
```

```
Requirement already satisfied: pillow>=9.0.0 in /usr/local/lib/python3.10/dist-
```

```
Requirement already satisfied: chardet in /usr/local/lib/python3.10/dist-packag
```

```
Installing collected packages: reportlab
```

```
Successfully installed reportlab-4.1.0
```

```
Collecting pdfminer.six
```

```
  Downloading pdfminer.six-20231228-py3-none-any.whl (5.6 MB)
```

```
5.6/5.6 MB 25.3 MB/s eta 0
```

```
Requirement already satisfied: charset-normalizer>=2.0.0 in /usr/local/lib/pyth
```

```
Requirement already satisfied: cryptography>=36.0.0 in /usr/local/lib/python3.1
```

```
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.10/dist-pac
```

```
Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-pack
```

```
Installing collected packages: pdfminer.six
```

```
Successfully installed pdfminer.six-20231228
```

```
from reportlab.lib.pagesizes import letter
from reportlab.pdfgen import canvas
```

```
def text_to_pdf(text, output_file):
    # Create a canvas with letter size
    c = canvas.Canvas(output_file, pagesize=letter)

    # Set up the font and size
    c.setFont("Helvetica", 12)

    # Split text into lines and write to the canvas
    lines = text.split('\n')
    y = 750 # Starting y position
    for line in lines:
        c.drawString(50, y, line)
        y -= 15 # Adjust vertical position for the next line

    # Save the canvas as a PDF file
    c.save()

# Example usage:
raw_text = ""
```

Chennai Super Kings (CSK) is an Indian professional cricket franchise based in Chennai, Tamil Nadu. The team was suspended for two years from the IPL starting July 2015 due to the involvement of the franchise in a betting scandal. BeacheseditMarina Beach is a 13-km-long urban beach along the Bay of Bengal that runs from the city center to the southern tip of the island.

Elliot's Beach, also known as Besant Nagar beach begins where the Marina ends. The beach is one of the most popular in Chennai. Chennai Rail Museum a railway museum in Perambur which has a rich rail heritage of India. Valluvar Kottam is a popular monument in Chennai, dedicated to the classical Tamil poet, P. H. Historic Government BuildingseditFort St. George is the name of the first British fortress in Chennai. It is considered as the largest mosque in the city of Chennai. ShoppingeditSpencer Plaza is one of the largest shopping malls in Chennai. Chennai has some unique places to offer for shopping. Art and crafts, contemporary and traditional. T. Nagar, the neighbourhood is the shopping hub of the city. Two main areas are Pondy Bazaar and T. Nagar.

There are four large amusement parks, MGM Dizzee World, VGP Universal Kingdom, Queen's Land

"""

```
output_file = "output.pdf"
text_to_pdf(raw_text, output_file)
```

```
from pdfminer.high_level import extract_text
text=extract_text('output.pdf')
print(repr(text))
print(text)
```

e-joined IPL for the 2018 season and won the title in its comeback season. The t

newspapers and periodicals published in India. The National Art Gallery building  
ly perforated hemispherical aluminium inner dome. It is located inside the Periy  
orld's 2nd oldest municipal corporation after the City of London Corporation. Th

```
import nltk
import numpy as np
import random
import string

nltk.download('punkt')
nltk.download('wordnet')
nltk.download('omw-1.4')
sentence_tokens=nltk.sent_tokenize(raw_text)
word_tokens=nltk.word_tokenize(raw_text)
lemmer=nltk.stem.WordNetLemmatizer()
def LemTokens(tokens):
    return [lemmer.lemmatize(token) for token in tokens]

remove_punct_dict=dict((ord(punct),None) for punct in string.punctuation)
def LemNormalize(text):
    return LemTokens(nltk.word_tokenize(text.lower().translate(remove_punct_dict)))

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...

greet_inputs=('hello','hi','whassup','namaste','kemchoo','how are you')
greet_responses=('hi','hey there',' i am fine','good','me too')
def greet(sentence):
    for word in sentence.split():
```

```

        if word.lower() in greet_inputs:
            return random.choice(greet_responses)

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity

def response(user_response):
    robot_response= ' '
    TfidfVec=TfidfVectorizer(tokenizer=LemNormalize,stop_words='english')
    tfidf=TfidfVec.fit_transform(sentence_tokens)
    vals=cosine_similarity(tfidf[-1],tfidf)
    idx=vals.argsort()[0][-2]
    flat=vals.flatten()
    flat.sort()
    req_tfidf=flat[-2]
    if(req_tfidf==0):
        robot_response=robot_response+ "I am sorry.Unable to understand!"
        return robot_response
    else:
        robot_response=robot_response+sentence_tokens[idx]
        return robot_response

flag = True
print("Hello I am retrieval learning chatbot please type your text after greeting to me fo
while(flag==True):
    user_response=input()
    user_response=user_response.lower()
    if(user_response!='bye'):
        if(user_response=='thank you' or user_response=='thanks'):
            flag=False
            print('Bot: You are welcome!')
        else:
            if(greet(user_response) !=None):
                print('Bot:' + greet(user_response))
            else:
                sentence_tokens.append(user_response)
                word_tokens=word_tokens+nlTK.word_tokenize(user_response)
                final_words=list(set(word_tokens))
                print('Bot:',end='')
                print(response(user_response))
                sentence_tokens.remove(user_response)
    else:
        flag = False
        print('Bot: Goodbye!')

Hello I am retrieval learning chatbot please type your text after greeting to me
Bot:/usr/local/lib/python3.10/dist-packages/sklearn/feature_extraction/text.py:
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/feature_extraction/text.py:409:
warnings.warn(
Two main areas are Pondy Bazaar and Ranganathan Street which are home to sever
There are four large amusement parks, MGM Dizzee World, VGP Universal Kingdom,
Bot: considered as the largest mosque in the city of Chennai.ShoppingeditSpence
Bot: Chennai has some unique places to offer for shopping.
Bot:

```

Chennai Super Kings (CSK) is an Indian professional cricket franchise based in Chennai, Tamil Nadu. It is one of the most successful franchises in the Indian Premier League (IPL), having won the tournament five times (2010, 2011, 2018, 2021, and 2023). The team is owned by the India Cements Limited and is captained by MS Dhoni.

Chennai Super Kings (CSK) is an Indian professional cricket franchise based in Chennai. The stadium is known for its vibrant atmosphere and is one of the most popular cricket grounds in India. The stadium is located in the heart of Chennai, near the Botanical Garden. Two main areas are Pondy Bazaar and Ranganathan Street which are home to some of the best shops and restaurants in the city. There are four large amusement parks, MGM Dizzee World, VGP Universal Kingdom, and the newly opened Chennai International Convention Centre.

--

Double-click (or enter) to edit